The Probabilistic Method

Week 11: Applications, Randomized Algorithms

Joshua Brody

CS49/Math59

Fall 2015

Reading Quiz

What resource is typically not an important factor in the analysis of streaming algorithms?

(A) memory space (hint: not the right answer)

(B) runtime

(C) power consumption

(D) approximation factor

(E) error guarantee

Reading Quiz

What resource is typically not an important factor in the analysis of streaming algorithms?

(A) memory space (hint: not the right answer)

(B) runtime

(C) power consumption

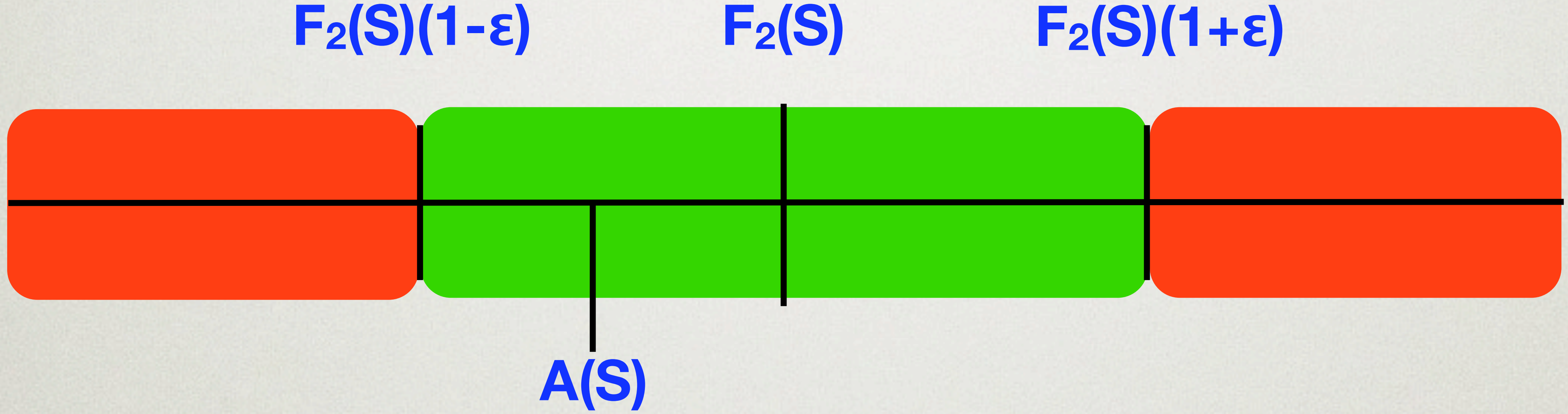
(D) approximation factor

(E) error guarantee

PAC Algorithms

Probably Approximately Correct (PAC) model:

"most of the time we're close enough"



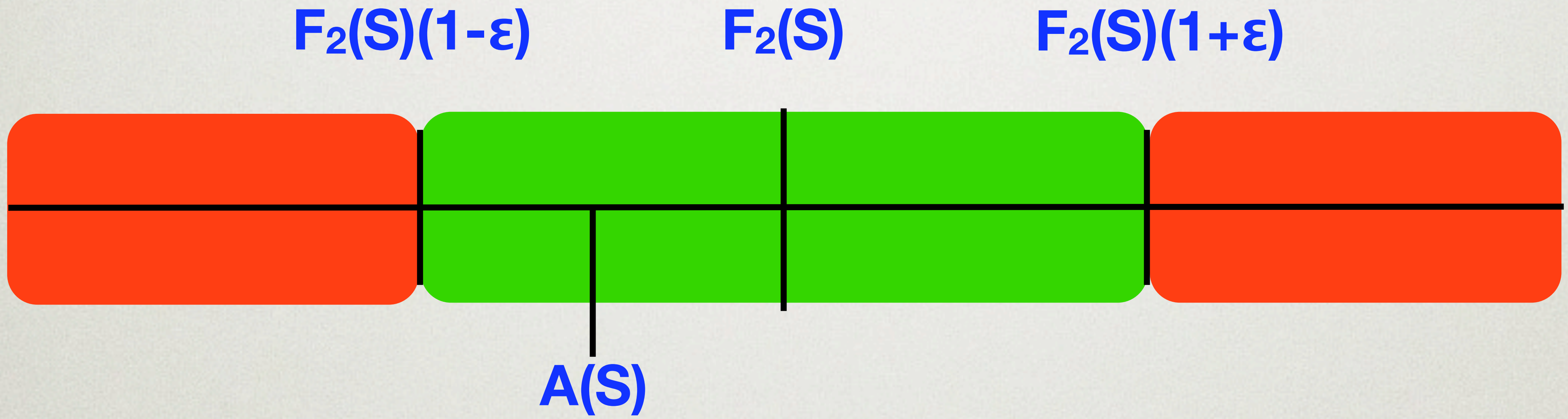
Definition: A streaming algorithm **A** is an (ε,δ)-approximation for **f** if for any stream **S**

$$\Pr[f(s)(1-\varepsilon) \leq A(S) \leq f(s)(1+\varepsilon)] \geq 1-\delta$$

PAC Algorithms

Probably Approximately Correct (PAC) model:

"most of the time we're close enough"



Theorem: For any $\varepsilon > 0$, there is a streaming algorithm that $(\varepsilon, 1/3)$-approximates for F_2 in $O([\log(n) + \log(m)]/\varepsilon^2)$ space.

Clicker Question

For what values of ε does the AMS Basic Estimator have error **< 1/3**?

(A) no ε makes this work

(B) ε > 1/3

(C) ε > π/2

(D) ε > √6

(E) None of the above

Clicker Question

For what values of ε does the AMS Basic Estimator have error **< 1/3**?

(A) **no ε makes this work**

(B) **ε > 1/3**

(C) **ε > π/2**

(D) **ε > √6**

(E) **None of the above**

Clicker Question

For arbitrary $\varepsilon > 0$, what values of k give error $< 1/3$?

(A) no k makes this work

(B) $k > 6/\varepsilon$

(C) $k > 6/\varepsilon^2$

(D) $k > \sqrt{6}$

(E) None of the above

Clicker Question

For arbitrary $\varepsilon > 0$, what values of **k** give error **< 1/3**?

(A) no k makes this work

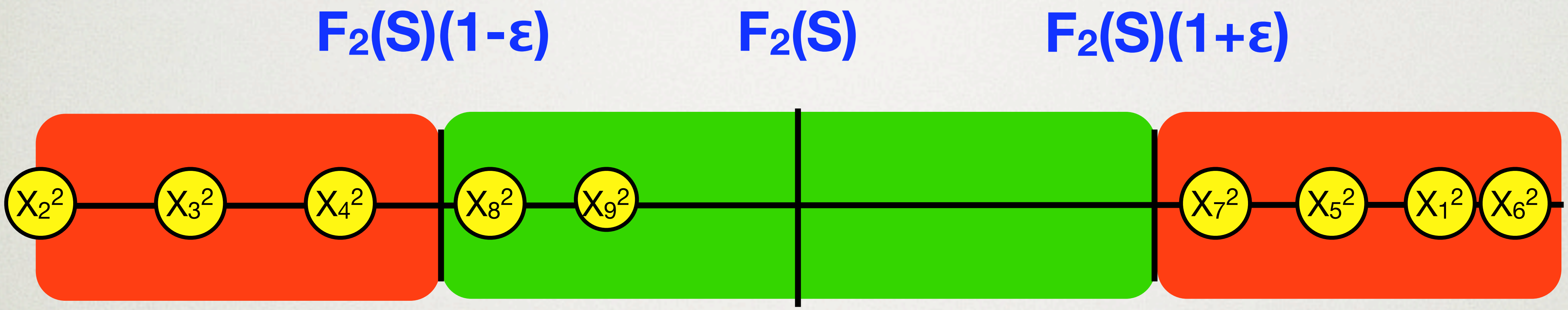
(B) $k > 6/\varepsilon$

(C) $k > 6/\varepsilon^2$

(D) $k > \sqrt{6}$

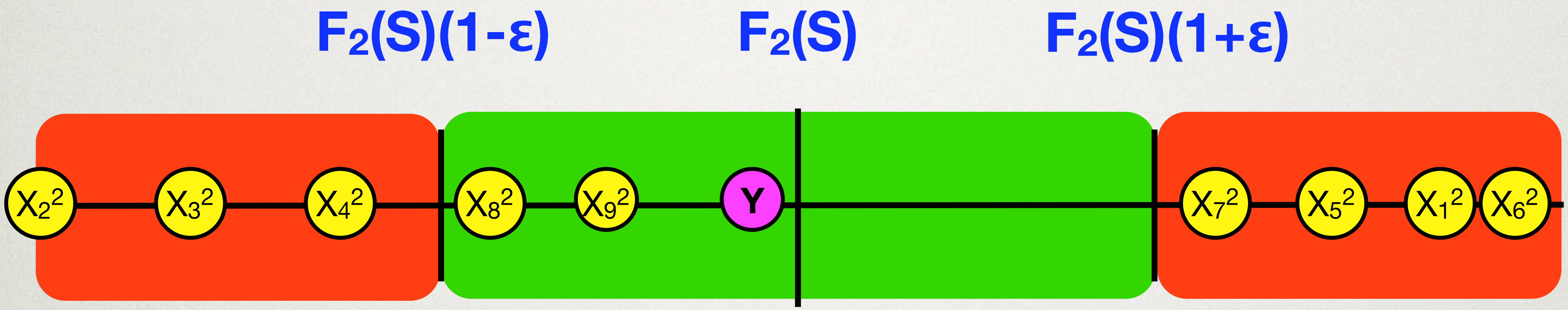
(E) None of the above

Improved F_2 Estimator



AMS Improved Estimator **$Y := (\sum X_i^2)/k$**

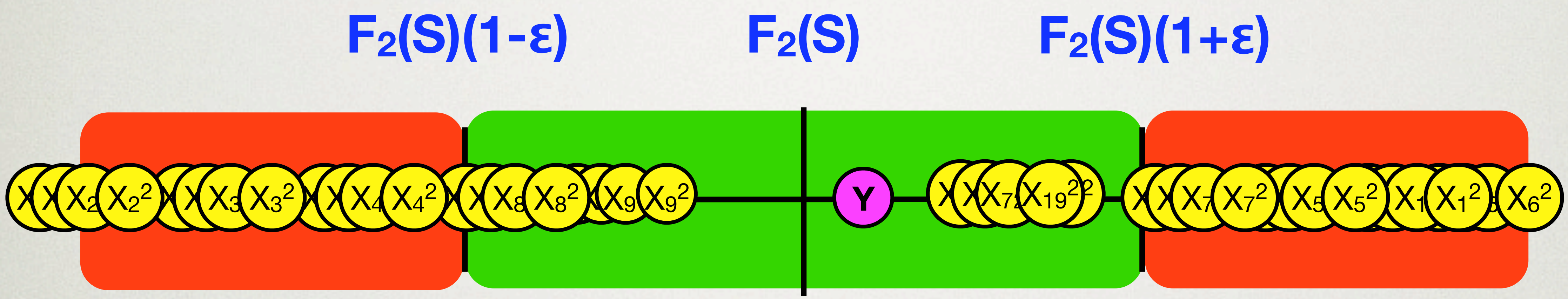
Improved F_2 Estimator



AMS Improved Estimator $Y := (\sum X_i^2)/k$

- $k = \Omega(1/\varepsilon^2)$: error $< 1/3$

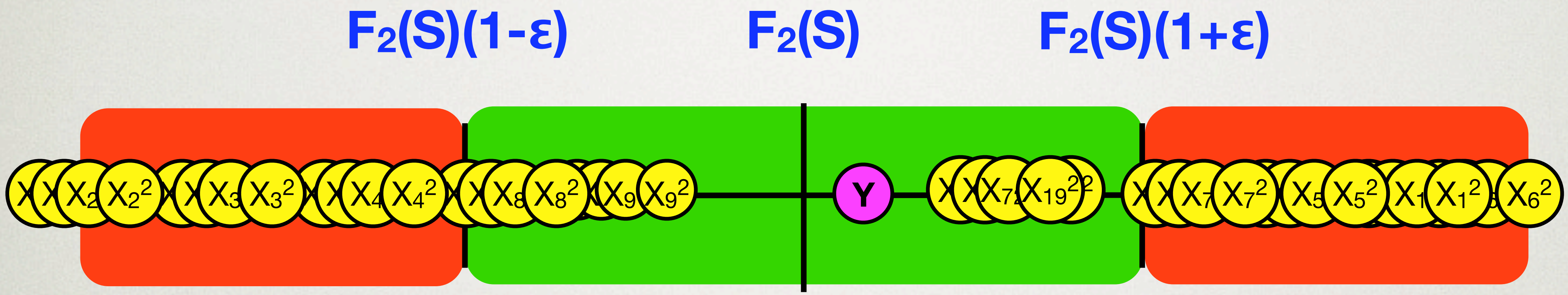
Improved F_2 Estimator



AMS Improved Estimator **Y:= $(\sum X_i^2)/k$**

- **$k = \Omega(1/\varepsilon^2)$** : error **< 1/3**
- **$k = \Omega(1/\delta\varepsilon^2)$** : error **< δ**

Improved F_2 Estimator

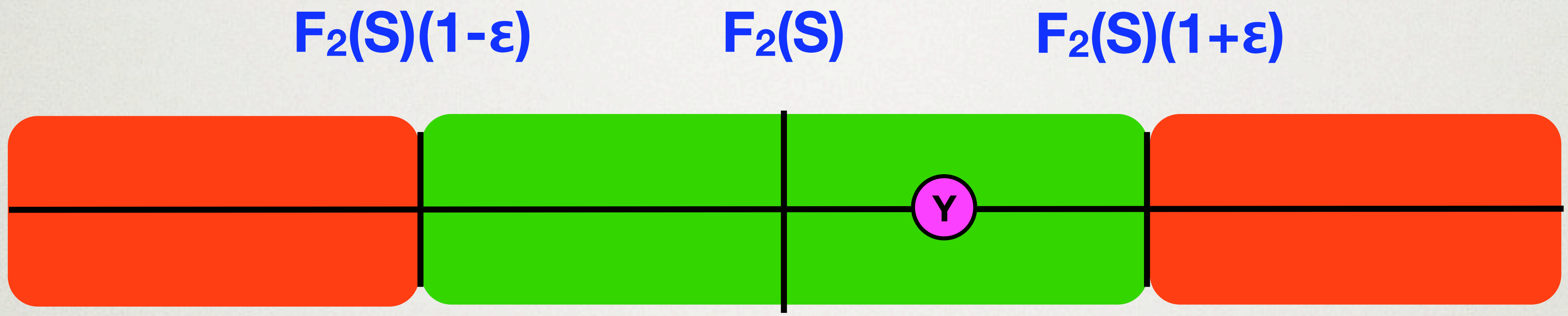


AMS Improved Estimator **$Y := (\sum X_i^2)/k$**

- **$k = \Omega(1/\varepsilon^2)$** : error **$< 1/3$**
- **$k = \Omega(1/\delta\varepsilon^2)$** : error **$< \delta$**

Better error dependence possible with Chernoff Bounds

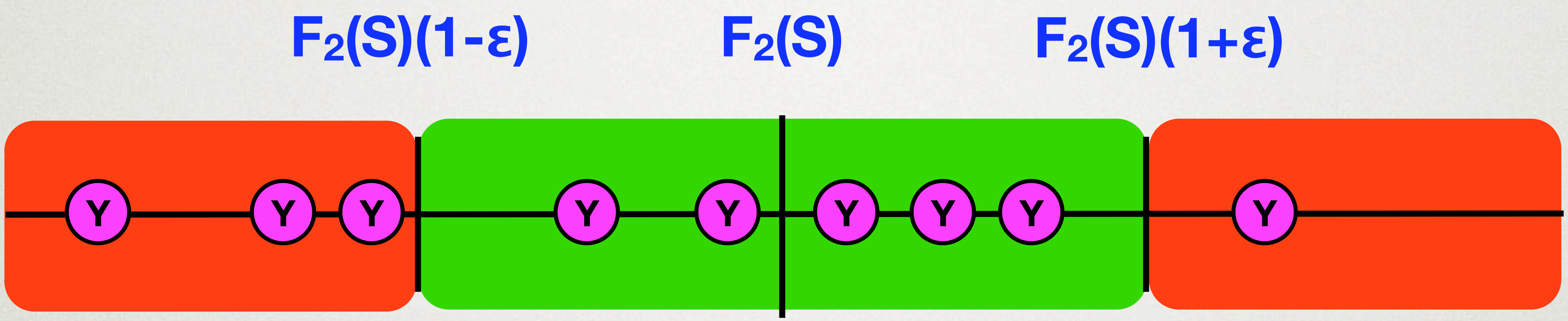
Final F_2 Estimator



AMS Improved Estimator $Y := (\sum X_i^2)/k$

- $k = \Omega(1/\varepsilon^2)$: error $< 1/3$

Final F_2 Estimator



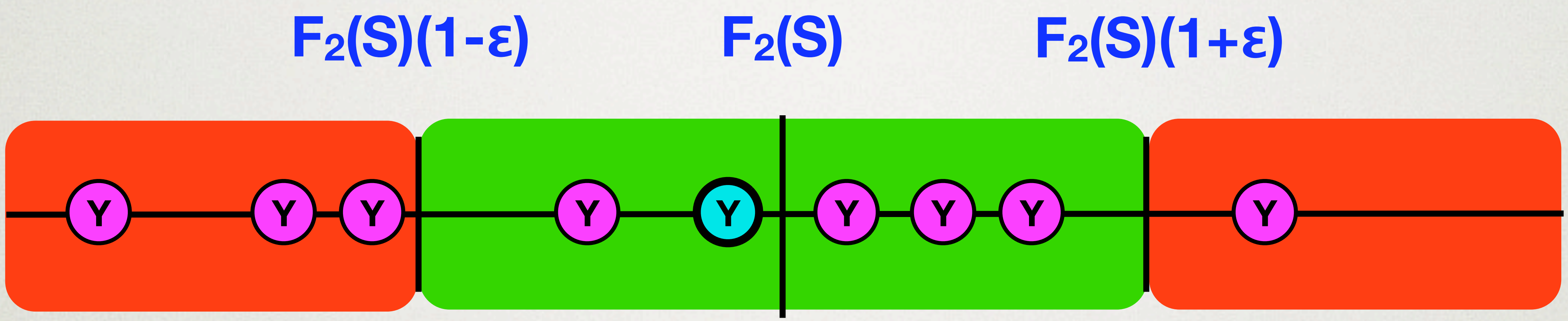
AMS Improved Estimator **$Y := (\sum X_i^2)/k$**

- **$k = \Omega(1/\varepsilon^2)$** : error **$< 1/3$**

AMS Final Estimator **median** of **k'** improved estimators

- **$k' = \Omega(\log(1/\delta))$** : error **$< \delta$**

Final F_2 Estimator



AMS Improved Estimator **$Y := (\sum X_i^2)/k$**

- **$k = \Omega(1/\varepsilon^2)$** : error **< 1/3**

AMS Final Estimator **median** of **k'** improved estimators

- **$k' = \Omega(\log(1/\delta))$** : error **< δ**

The Probabilistic Method